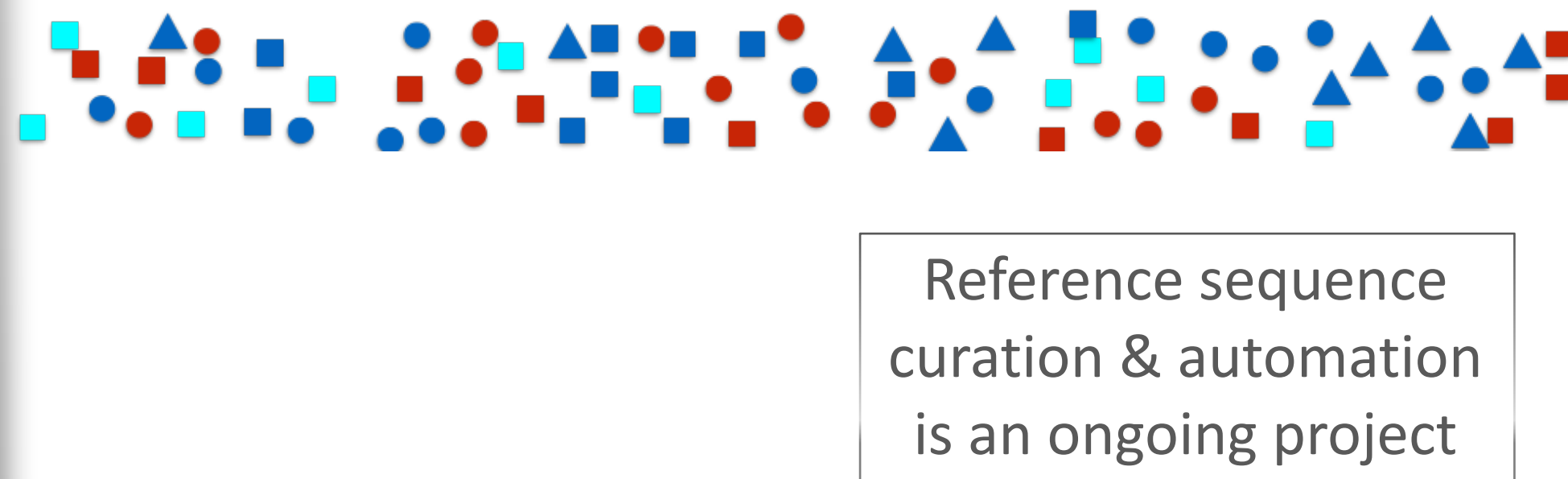


Background

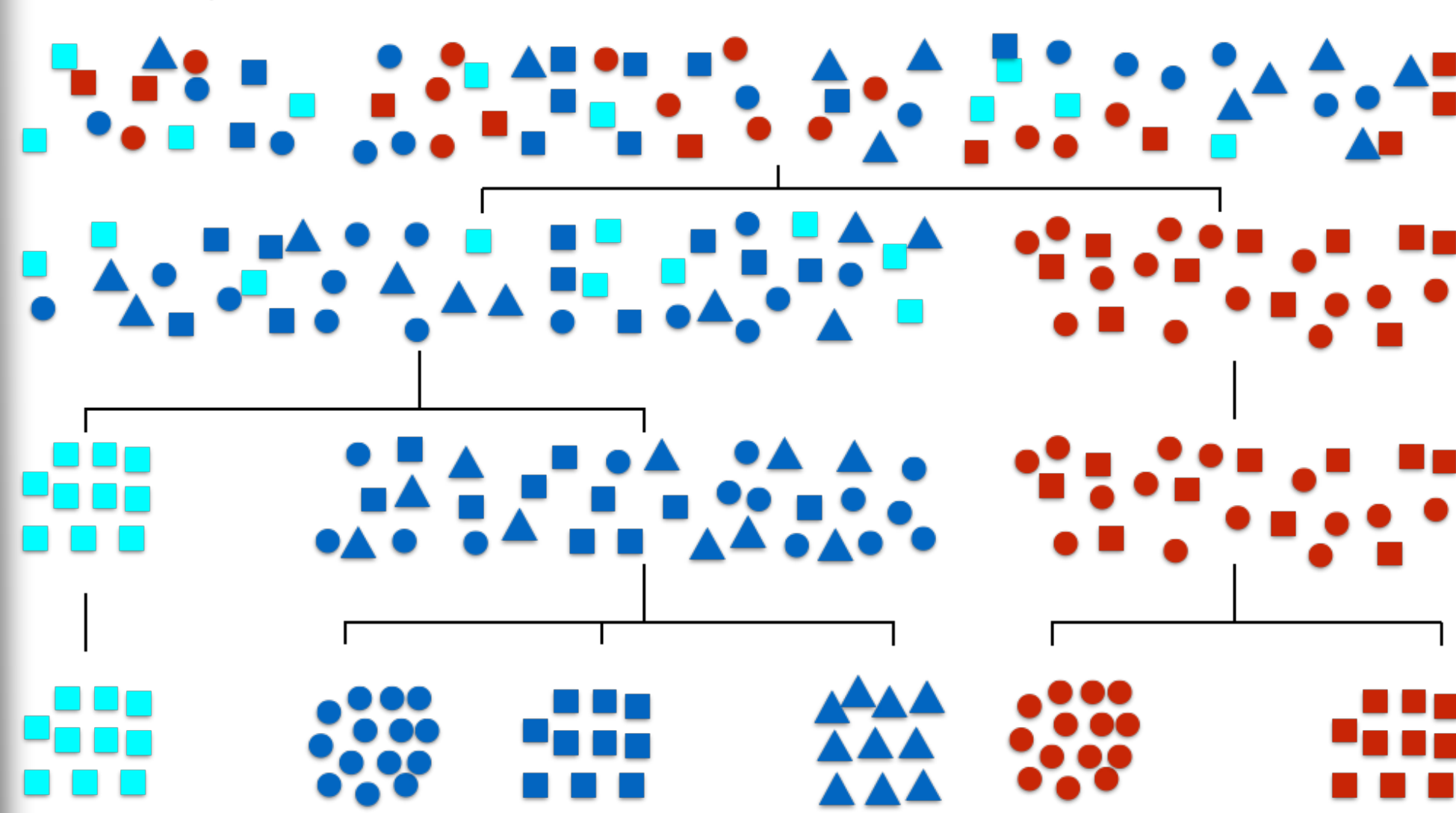
Clustering of sequences into Operational Taxonomic Units (OTUs) has become a mainstream approach to facilitate taxonomic classification of large numbers of 16S rRNA gene sequences. This is partly due to the high computational requirements for processing each sequence in increasingly large datasets. A primary focus of the field has been development and improvement of OTU-based sequence clustering methods that rely on distances between each pair of sequences in a dataset. Following OTU-based clustering, representative sequences are commonly classified using tools such as the RDP Naïve Bayesian Classifier, and the resulting classification transitively assigned to all sequences comprising that OTU. However, problems with this strategy exist¹. Here, we present PECAN, a novel per sequence taxonomic assigner which quickly and accurately classifies millions of 16S rRNA gene sequences. MC Classifier relies on higher order Markov Chain models built from a user-specified set of reference sequences. These models are used to estimate the probability that a query sequence belongs to a particular taxonomic rank. We show that this method requires less computational resources than other algorithms, removes the necessity and issues associated with OTU-based sequence clustering, and is applicable to any microbiome dataset.

Methods & Usage

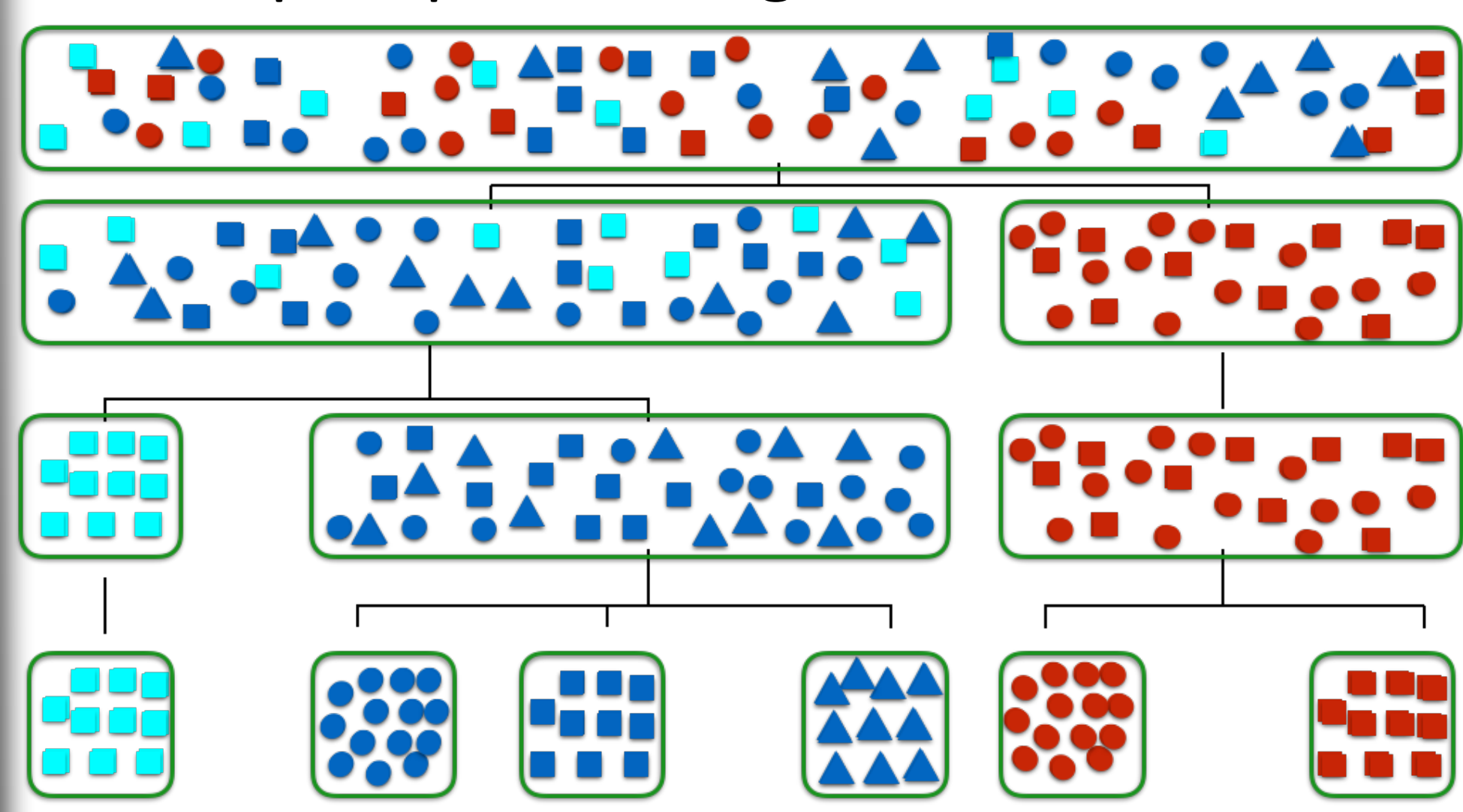
1. Collect reference sequences



2. Align sequences & build taxonomic tree

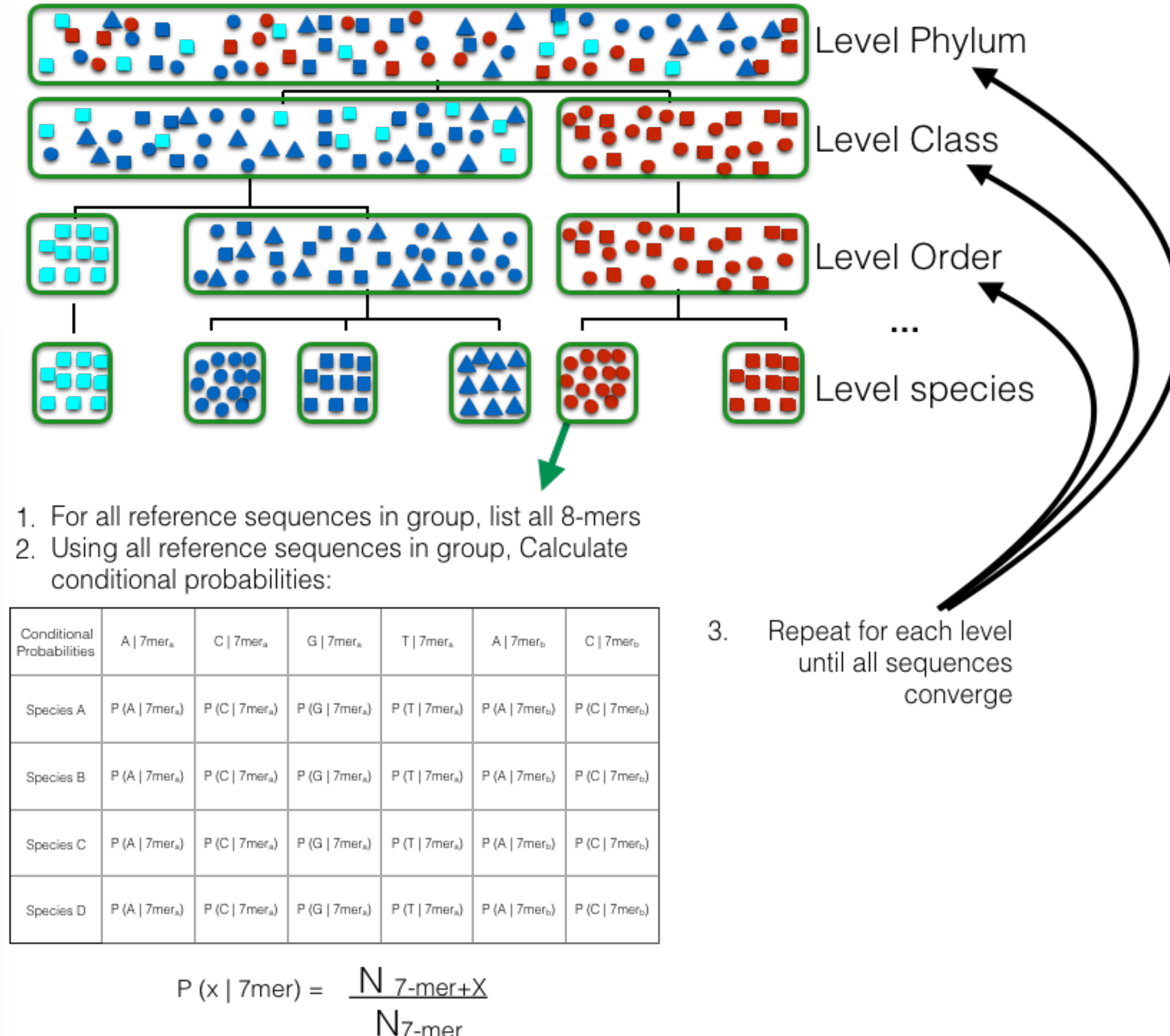


3. Group sequences together at each level

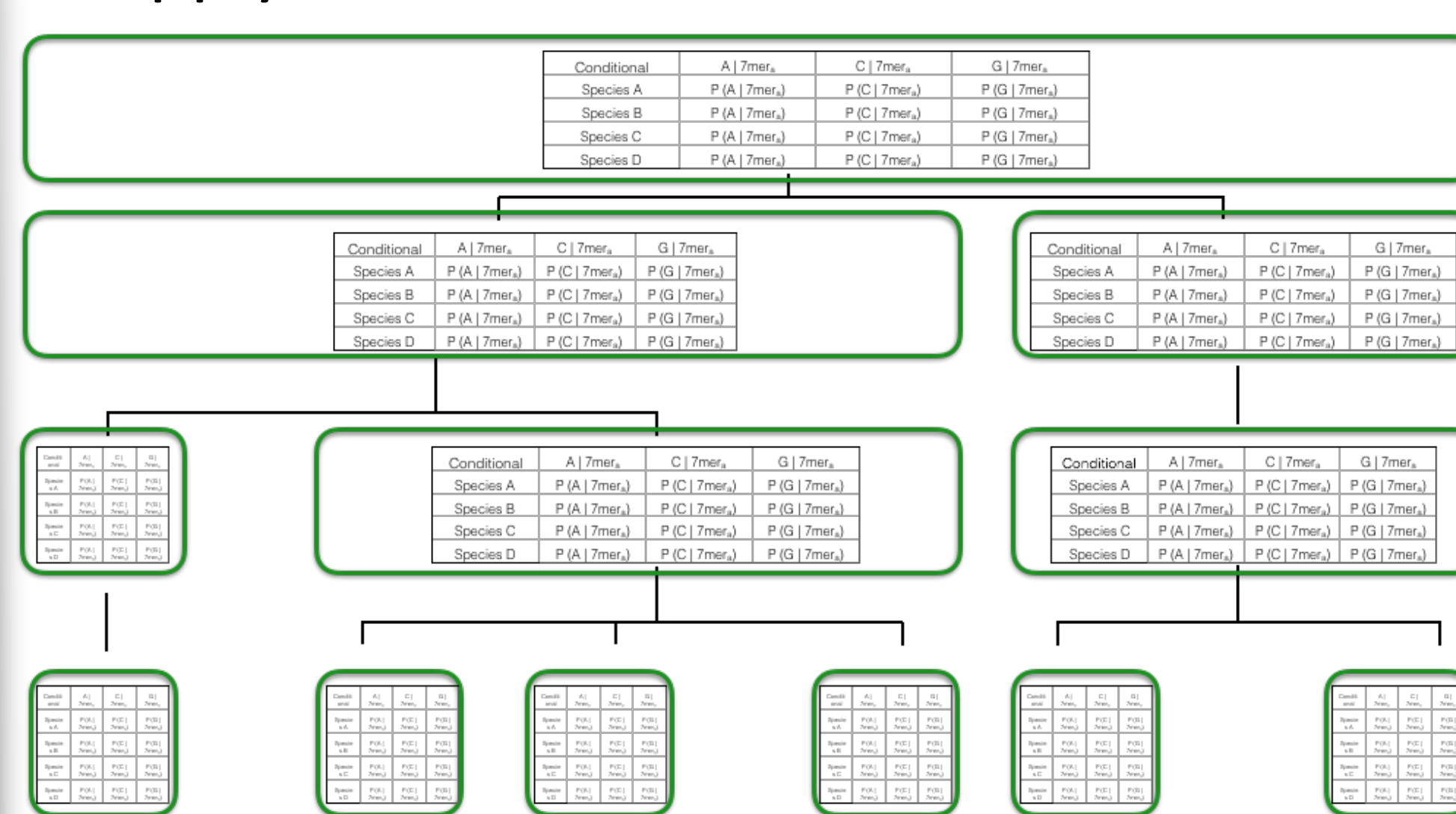


Methods & Usage (cont.)

4. Build Markov Chain models

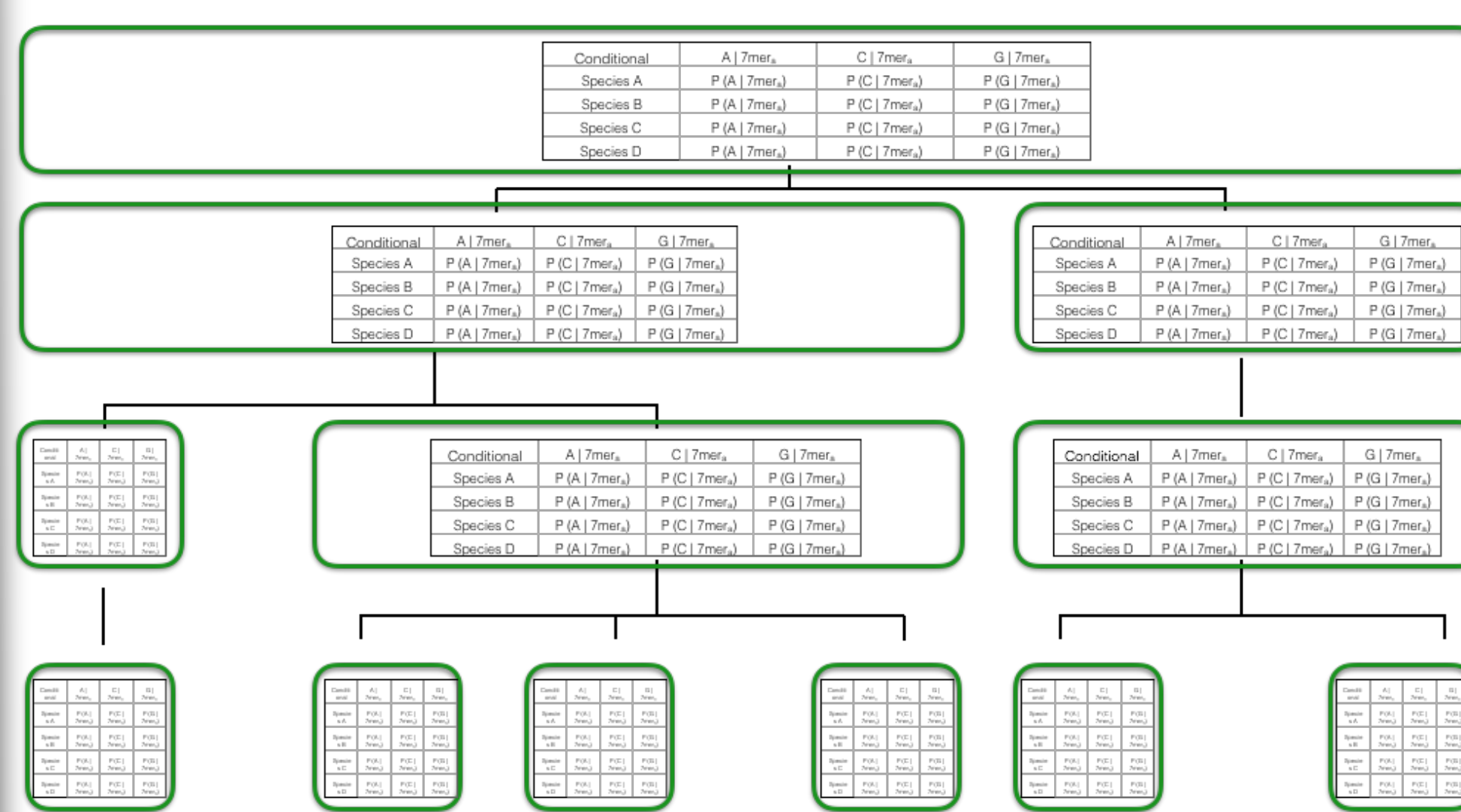


4. Apply models to taxonomic tree



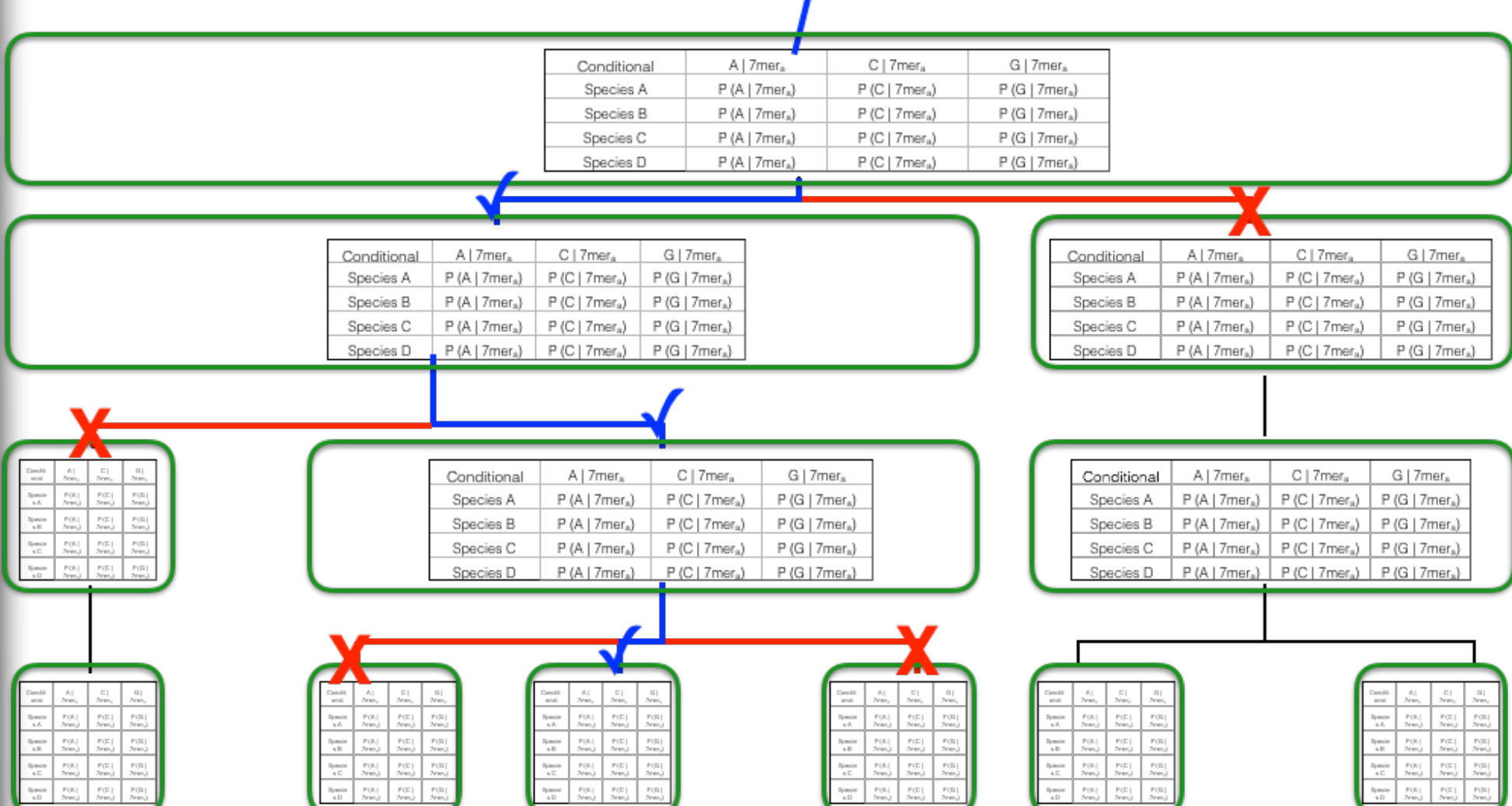
5. Usage: Apply query sequence

■ Query Sequence



6. Usage: Calculate posterior probability of query sequence at each level

■ Query Sequence



Results

Results

References

Acknowledgements